Work Report-Unsupervised Recommendation Annotating Model

Work Report-Unsupervised Recommendation Annotating Model

- Problem review
- The application of Variational autoencoder model
  - Basic assumption[1]
    - Manifold Hypothesis
    - VAE hypothesis based on manifold hypothesis
  - Loss function
    - Loss function for VAE
    - Loss function for Unsupervised Recommendation Annotating Model
  - Model structure
- Our innovation points
  - Pretraining strategy
  - Unsupervised recommendation annotation method
  - Adapting VAE to medical image analysis tasks
- The next working arrangement

Reference

Problem review

In many biomedical processing tasks, simply increasing the amount of annotated data can't improve the performance of deep model, for the extremely large similarity between biomedical images and the highly coincident features in them. As shown in table,the deep learning model can achieve respectable performance with partial training data(e.g. random select 30%,40%,50% from full dataset), which indicates that the accuracy of the model does not increase significantly compared with the increasing of annotated data number. This means that a large quantity of the annotated data is wasted and they do not contribute new features to the model. However, the annotation cost for biomedical image is extremely high. How to filter unlabeled images and find representative data with different features under limited annotation cost is a very important problem.

Medical Image Processing Task	Random Select Percentage	Random Result	Full Result
lung nodule segmentation (Dice)	42%	73%	74.4%
lung nodule classification (Accuracy)	42%	85%	87%
Lymph nodule segmentation (Mean IU)	50%	85.8%	87.9%
Gland segmentation (Dice)	10%	82%	90%
	30%	86%	
	50%	87%	

We state our problem qualitatively as this :(we will give the math formulation of this problem under manifold hypothesis and latent variable hypothesis)

Suppose we have a set of biomedical images $X = \{X_1, X_2, \ldots, X_n\}$, and we want to choose the most representative k images to annotate, we want the k images cover the feature space of I as much as possible.

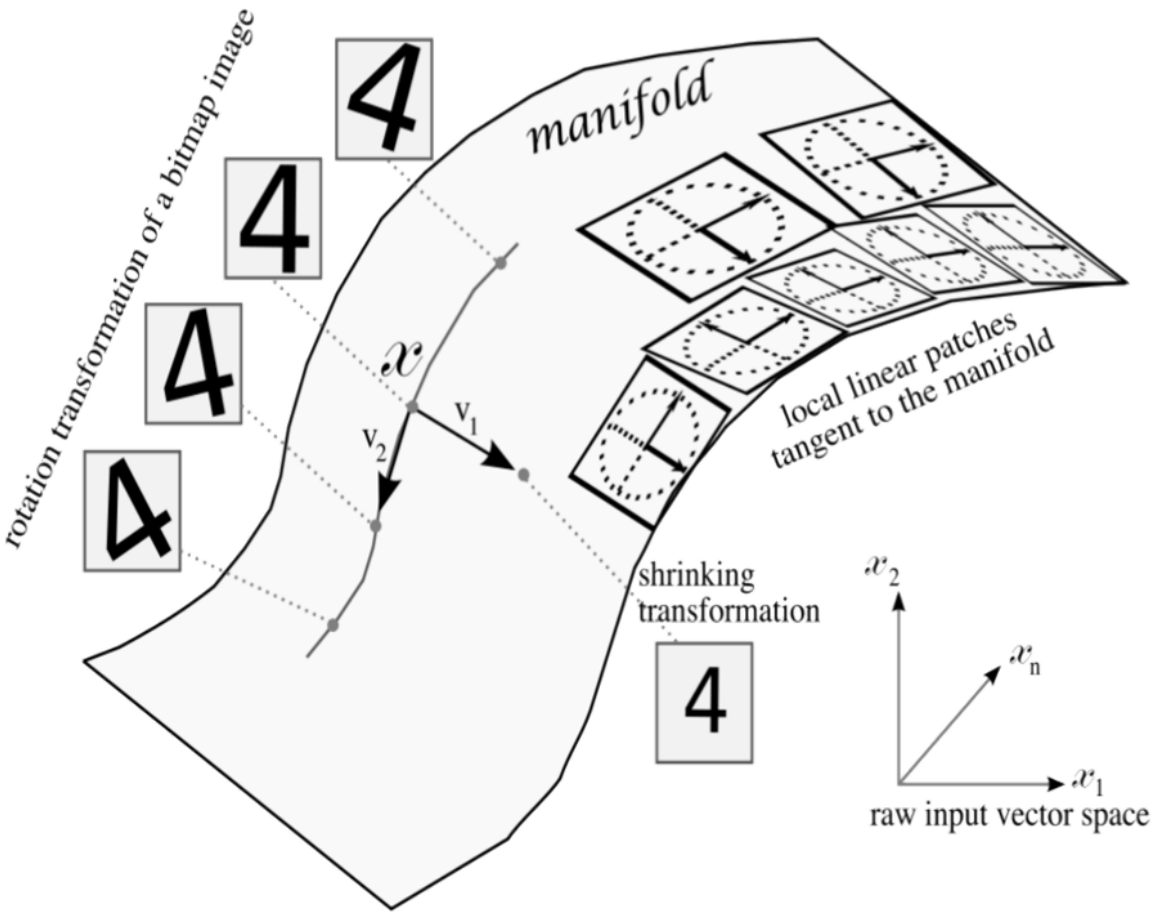
The application of Variational autoencoder model

Basic assumption[1]

Manifold Hypothesis

Suppose images $\{X_1, X_2, \ldots, X_n\}$ are points in high dimensional space $\mathcal{X}$, and this points are all lying on a manifold embedding in $\mathcal{X}$, which is generated by low dimensional latent space $\mathcal{Z}$. And there exists a family of continuous deterministic functions f, parameterized by a vector θ_f in some space Θ_f, where $f : \mathcal{Z} \times \Theta_f \rightarrow \mathcal{X}$. For each point $X \in \mathcal{X}$, we have at least one point $z \in \mathcal{Z}$, satisfy $X = f(z, \theta_f)$.

The manifold Hypothesis can be visualized as this :



VAE hypothesis based on manifold hypothesis

Suppose high dimensional space $\mathcal{X}$ is a probability space with probability density function P, and images $\{X_1, X_2, \ldots, X_n\}$ are sampled from P.(1)

Suppose the latent variable space $\mathcal{Z}$ is also a probability space with prior distribution $P(z) \sim N(0, I)$. (2)

Suppose we have a family of continuous deterministic functions f, parameterized by a vector θ_f in some space Θ_f, f generate points in $\mathcal{X}$ with samples in $\mathcal{Z}$:

$$f : \mathcal{Z} \times \Theta_f \to \mathcal{X}$$

If we see $z \in \mathcal{Z}$ is random and $\theta_f \in \Theta_f$ is fixed, we assume that

$$P(X|z, \theta_f) = N(X|f(z, \theta_f), \sigma^2 * I), z \in \mathcal{Z}, \theta_f \in \Theta_f, c$$

So that we can calculate $P(X)$ with:

$$P(X) = \int_z P(X|z; \theta) P(z) dz$$

In VAE we assume the possible distribution of latent space which most possibly generate X is a gaussian distribution with:

$$Q(z|X) \sim N(z|u, \Sigma)$$

Suppose we also have another family of continuous deterministic function g, parameterized by a vector θ_g in some space Θ_g, g map $X \in \mathcal{X}$ into one possible distribution of latent space which most possibly generate X:

$$g : \mathcal{X} \times \Theta_g \rightarrow \mu \times \Sigma$$

Loss function

Loss function for VAE

We want to maximize the log likelihood of $log(P(X))$ and we want to minimize the difference between $Q(z|X)$ and the real conditional distribution of z under X annotated with $P(z|X)$, so the loss function for VAE is:

$$log(P(X)) - \mathcal{D}[Q(z|X)||P(z|X)]$$

We transform it into :

$$E_{z\sim Q(z|X)}[log(P(X|z))] - \mathcal{D}[Q(z|X)||P(z)]$$

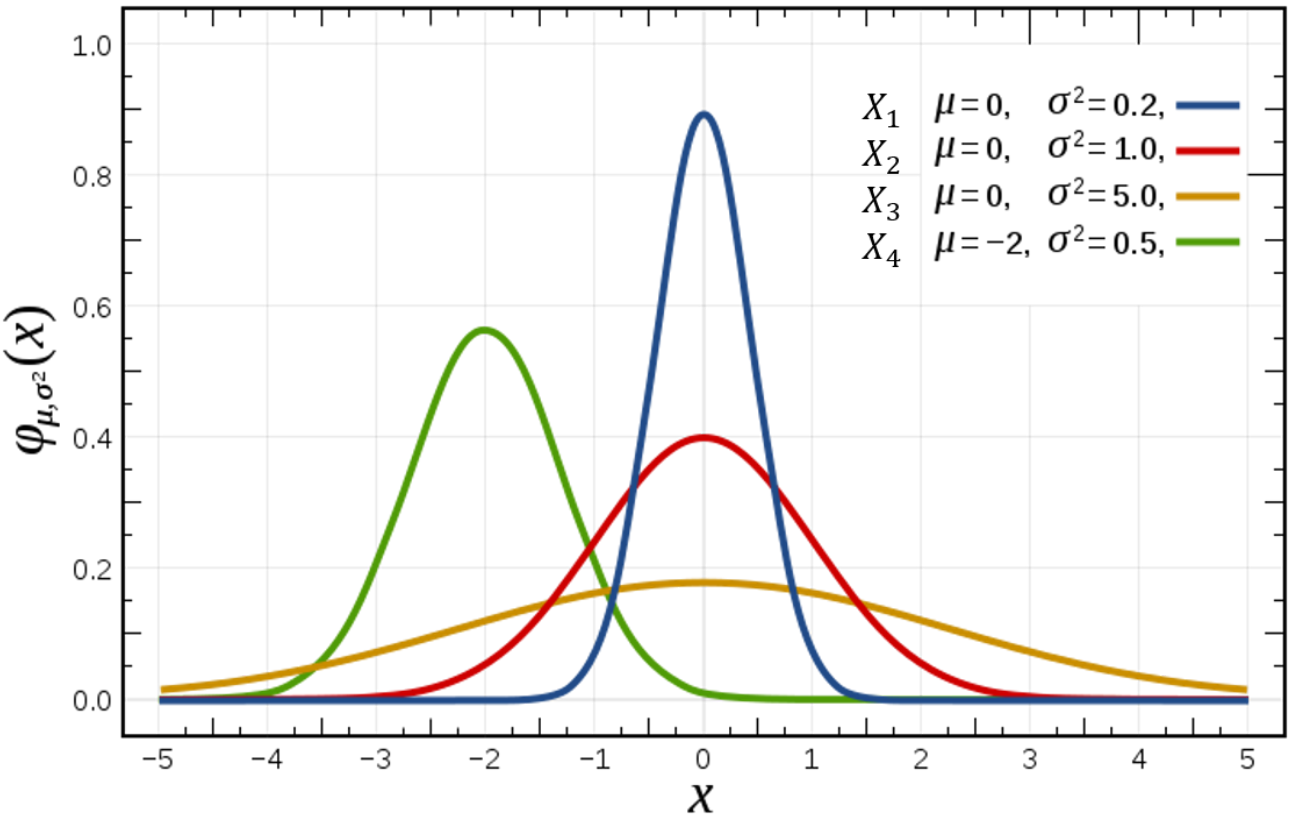
Loss function for Unsupervised Recommendation Annotating Model

By using gradient descent to optimize (7), we get the conditional latent variable distribution for each $X_i \in X$。 the distribution of $Q(z|X_i) = N(\mu_i, \Sigma_i)$ give us an intuition about which is the most possible z for generating X_i (the specific latent variable distribution for X_i). Review our target:

> *Suppose we have a set of biomedical images $X = \{X_1, X_2, \ldots, X_n\}$, and we want to choose the most representative k images to annotate, we want the k images cover the feature space of I as much as possible.*

A natural thought is like this:

Suppose the conditional latent variable distribution for X_1, X_2, X_3, X_4 is plot:



If we want to choose 2 images to represent X_1, X_2, X_3, X_4, we may choose (X_1, X_4) rather than X_1, X_2, for (X_1, X_4) can cover more area in latent space than (X_1, X_2),and the latent variable for generate X_1 is much more likely able to generate X_2 than X_4.

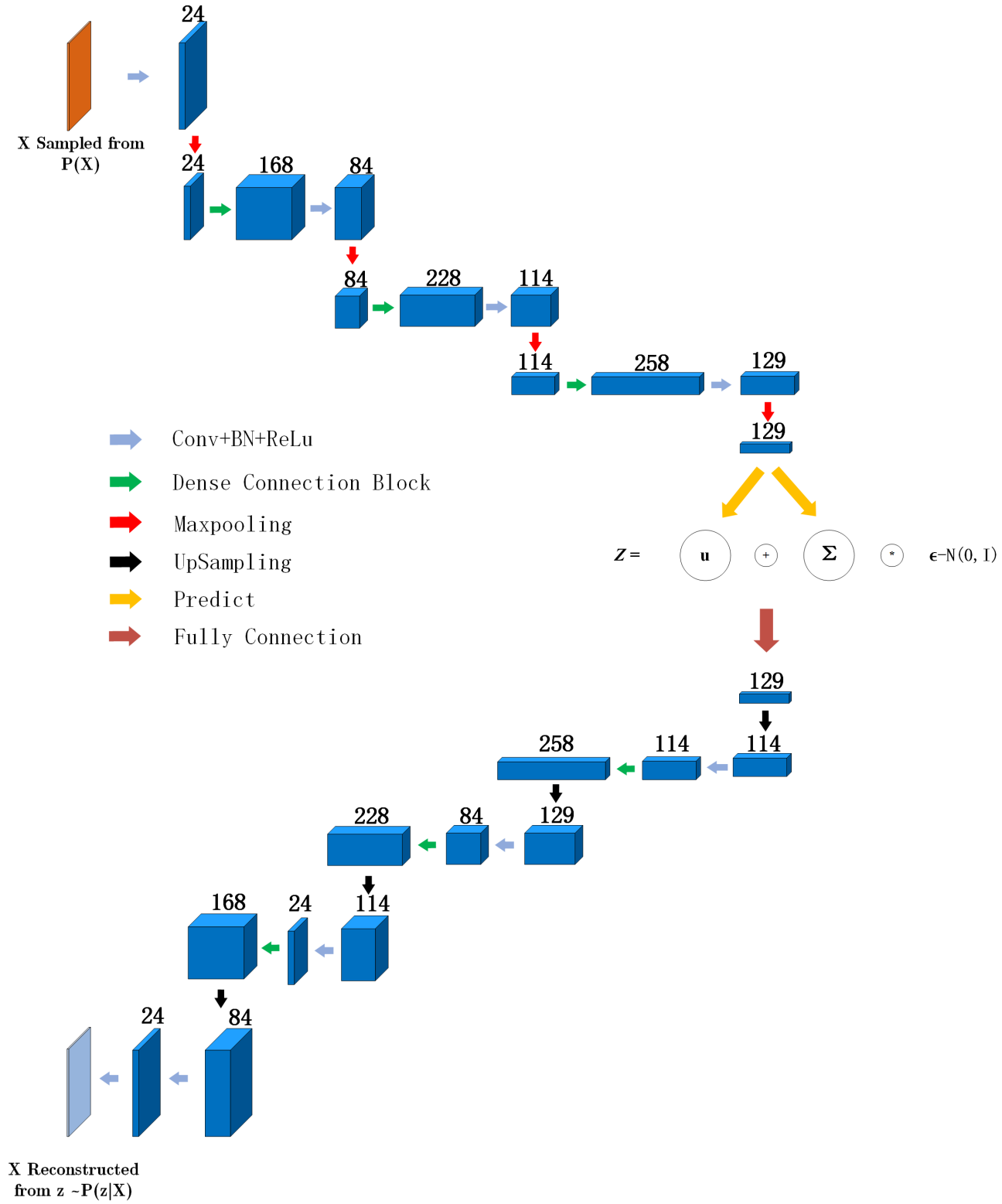
Based on this intuition ,we can form the target into this form:

> *Suppose we have a set of biomedical images* $X = \{X_1, X_2, \ldots, X_n\}$*, and have their conditional latent variable distribution* $N(\mu_1, \Sigma_1), \ldots, N(u_n, \Sigma_n)$ *.we want to choose* k *images to annotate, the aim is to maximize:*

$$\sum_{i=1}^{k}\sum_{j=1}^{k}\mathcal{D}[N(\mu_i,\Sigma_i)\|N(u_j,\Sigma_j)]$$

(Note : It's one of the loss functions we want to try)

Model structure



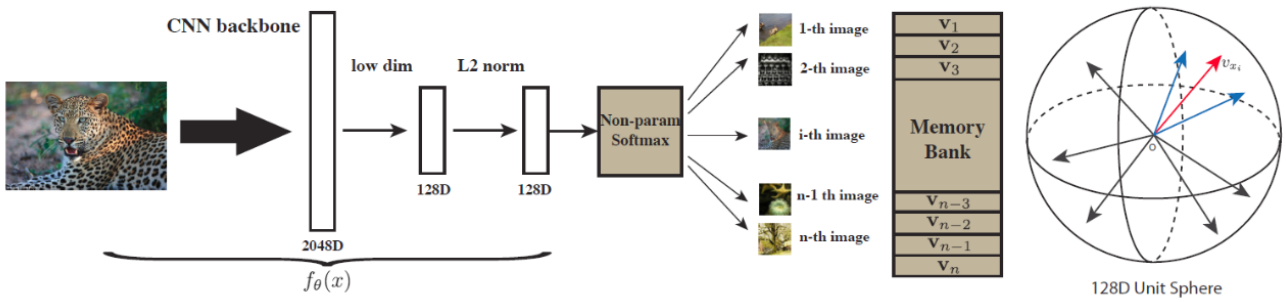
Here we want to use the structure in our last paper to build the main structure model and use VAE to predict the distribution of latent variable.The loss function of our model is:

$$loss = \|X - X_{Recon}\|^2/\sigma^2 - [\|u\|_2^2 + \sum_i \Sigma_{i,i}^2 - \sum_i ln(\Sigma_{i,i}^2) -$$

Our innovation points

Pretraining strategy

We can use the method in [2] to do pretraining to make the distribution of latent variable more discriminative.



Unsupervised recommendation annotation method

The previous work about recommendation annotation are all based of adaptive learning framework, which is an supervised method and requires a good model to process specific tasks.

Our method is a totally unsupervised method, which is an innovation point.

Adapting VAE to medical image analysis tasks

There have been only 2 papers applying VAE to MIA tasks[3], [4], one in Retinal Fundus Images and the other in MRI images. So our method proposes a new direction to utilize VAE. We use VAE to illustrate the math assumption of recommendation annotation ,which makes the problem more clear. We also use distribution assumption to define "representative", which can help us design loss function(e.g. (8)).

The next working arrangement

1. Finish the autumn semester coursework and exams
2. Try our idea in gland segmentation dataset and lung nodule classification data set
3. Reading as much papers as possible about the application of VAE or related researches, and figure out how to set network structure, how to make different distribution hypothesis. The main application of VAE is in face recognition field.

Reference

[1] Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[2] Wu Z, Xiong Y, Stella X Y, et al. Unsupervised Feature Learning via Non-Parametric Instance Discrimination[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 3733-3742.

[3] Sedai S, Mahapatra D, Hewavitharanage S, et al. Semi-supervised Segmentation of Optic Cup in Retinal Fundus Images Using Variational Autoencoder[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2017: 75-82.

[4] Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data[J]. Magnetic resonance in medicine, 2018, 79(6): 3055-3071.